Application of Information Theory, Lecture 10

Hardcore Predicates

Handout Mode

Iftach Haitner

Tel Aviv University.

December 22, 2015

Part I

Motivation and Definition

Hardcore predicates

- Let $f\colon \{0,1\}^n \mapsto \{0,1\}^n$ be a “hard to invert" function, how unpredictable is x given $f(x)$
- Parts of x might be (totally) predictable
- It turns out that there is an hardcore part in x.

Hardcore predicates, cont.

Definition 1 (hardcore predicates)

A predicate $b\colon \{0,1\}^n \mapsto \{0,1\}$ is (s,ε)-hardcore predicate of $f\colon \{0,1\}^n \mapsto \{0,1\}^n$, if $\Pr_{x\leftarrow\{0,1\}^n}[\mathsf{P}(f(x)) = b(x)] \leq \frac{1}{2} + \varepsilon$, for any s-size P.

- Why size?
- We will typically consider poly-time computable f and b.
- Does every function has such a predicate?
- Does every hard to invert function has such a predicate?
- Is there a generic hardcore predicate for all hard to invert functions?

  Let f be a function and let b be a predicate, then b is typically not a hard-core predicate of $g(x) = (f(x), b(x))$.

Part II

The Information Theoretic Settings

Some definitions

Let $f\colon \mathcal{D} \mapsto \mathcal{R}$.

- $\mathrm{Im}(f) = \{f(x)\colon x \in \mathcal{D}\}$.
- $f^{-1}(y) = \{x \in \mathcal{D}\colon f(x) = y\}$
- f is d regular, if $\left|f^{-1}(y)\right| = d$ for every $y \in \mathrm{Im}(f)$.
- min entropy of $X \sim p$ is
  $\mathrm{H}_\infty(X) = \min_{x\in\mathcal{X}}\{-\log p(x)\} = -\log \max_{x\in\mathcal{X}}\{p(x)\}$.
- Examples:
  - Z is uniform over 2^k-size set.
  - $Z = X\mid_{f(X)=y}$, for 2^k-regular f, $y \in \mathrm{Im}(f)$ and $X \leftarrow \mathcal{D}$.
- In both examples $\mathrm{H}_\infty(Z) = k$

2-universal families

Definition 2 (2-universal families)

A function family $\mathcal{G} = \{g \colon \mathcal{D} \mapsto \mathcal{R}\}$ is 2-universal, if $\forall\, x \neq x' \in \mathcal{D}$ it holds that $\Pr_{g \leftarrow \mathcal{G}}\left[g(x) = g(x')\right] = \frac{1}{|\mathcal{R}|}$.

Example: $\mathcal{D} = \{0,1\}^n$, $\mathcal{R} = \{0,1\}^m$ and $\mathcal{G} = \{A \in \{0,1\}^{m \times n}\}$ with $A(x) = A \times x \bmod 2$.

Lemma 3 (leftover hash lemma)

Let X be a rv over $\{0,1\}^n$ with $\mathrm{H}_2(X) \geq k$ let $\mathcal{G} = \{g \colon \{0,1\}^n \mapsto \{0,1\}^m\}$ be 2-universal and let $G \leftarrow \mathcal{G}$. Then $\mathrm{SD}((G, G(X)), (G, \sim \{0,1\}^m)) \leq \frac{1}{2} \cdot 2^{(m-k)/2}$.

Hardcore predicate for regular functions

Lemma 4

Let $f\colon \{0,1\}^n \mapsto \{0,1\}^n$ be 2^k-regular function, let $\mathcal{G} = \{g\colon \{0,1\}^n \mapsto \{0,1\}\}$ be 2-universal and let $v\colon \{0,1\}^n \times \mathcal{G} \mapsto \{0,1\}^n \times \mathcal{G}$ be defined by $v(x,g) = (f(x), g)$.
Then $b(x,g) = g(x)$ is $(\infty, 2^{-(k-1)/2})$ hardcore-predicated of v.

- b is an hardcore predicate of v (not of f)

Proving Lemma 4

Claim 5

$\mathrm{SD}\left((f(X), G, G(X)), (f(X), G, U)\right) \leq 2^{-(k-1)/2}$,
for $G \leftarrow \mathcal{G}$, $X \leftarrow \{0,1\}^n$ and $U \leftarrow \{0,1\}$.

We conclude the proof showing that indistinguishability implies unpredictability.

Lemma 6 (predicting to distinguishing)

Let (Y, Z) be rv over $\{0,1\}^* \times \{0,1\}$ and let P be an algorithm with $\Pr[\mathsf{P}(Y) = Z] \geq \frac{1}{2} + \varepsilon$. Then $\exists$ algorithm D, with essentially the same complexity as P, with $\Pr[\mathsf{D}(Y, Z) = 1] - \Pr[\mathsf{D}(Y, U) = 1] \geq \varepsilon$.

Proof: $\mathsf{D}(y, z)$ outputs 1 if $\mathsf{P}(y) = z$ and 0 otherwise.$\square$

Corollary 7

If $\mathrm{SD}((Y, Z), (Y, U)) < \varepsilon$, then $\Pr[\mathsf{P}(Y) = Z] < \frac{1}{2} + \varepsilon$ for any predictor P.

Proving Claim 5

For $y \in \text{Im}(f)$, let X_y be uniformly distributed over $f^{-1}(y)$.
Compute

$$
\begin{aligned}
&\text{SD}((f(X), G, G(X)), (f(X), G, U)) \\
&= \sum_{y \in \text{Im}(f)} \Pr[f(X) = y] \cdot \text{SD}\big((y, G, G(X)|_{f(X)=y}), (y, G, U)\big) \quad \text{(board)} \\
&= \sum_{y \in \text{Im}(f)} \Pr[f(X) = y] \cdot \text{SD}\left((y, G, G(X_y)), (y, G, U)\right) \\
&\leq \max_{y \in \text{Im}(f)} \text{SD}((y, G, G(X_y)), (y, G, U)) \\
&= \max_{y \in \text{Im}(f)} \text{SD}((G, G(X_y)), (G, U))
\end{aligned}
$$

Since $\text{H}_\infty(X_y) = k$ for every $y \in Im(f)$, the leftover hash lemma yields that

$$
\begin{aligned}
\text{SD}((G, G(X_y)), (G, U)) \leq& \frac{1}{2} \cdot 2^{(1-\text{H}_\infty(X_y)))} \\
&= 2^{(-k-1)/2}. \square
\end{aligned}
$$

Part III

The Computational Settings

Hard functions

An injective function has hardcore bit, only if it is "hard to invert".

Definition 8 (hard function)

$f\colon \{0,1\}^n \mapsto \{0,1\}^n$ is (s,ε)-hard, if
$\Pr_{x \leftarrow \{0,1\}^n}\left[\mathsf{Inv}(f(x)) \in f^{-1}(f(x))\right] \leq \varepsilon$ for any s-size Inv.

- Size? Length preserving?
- f is hard $\implies$ predicting x from $f(x)$ is hard.
- But does any hard function has an hardcore predicate?
- f is injective and not hard $\implies$ f has no hardcore predicate.

The Goldreich-Levin predicate

For $x, r \in \{0,1\}^n$, let $\langle x, r\rangle_2 := (\sum_{i=1}^n x_i \cdot r_i) \bmod 2 = \bigoplus_{i=1}^n x_i \cdot r_i$.

Theorem 9 (Goldreich-Levin)

For $f\colon \{0,1\}^n \mapsto \{0,1\}^n$, define $g\colon \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}^n \times \{0,1\}^n$ by $g(x,r) = (f(x), r)$. Assume f is (s, ε)-hard, then $b(x,r) := \langle x, r\rangle_2$ is an $(\frac{\varepsilon}{n^2} \cdot s, \sqrt[3]{n\varepsilon})$-hardcore predicate of g.

- Parameters are not tight, and we ignore small terms.
- If f is $(n^{\omega(1)}, 1/n^{\omega(1)})$-hard, then b is an $(n^{\omega(1)}, 1/n^{\omega(1)})$-hardcore predicate of g.
- Proof by reduction: a too small P for predicting $b(x,r)$ "too well" from $(f(x), r)$, implies a too small inverter for f:
- Assume $\exists\ s'$-size P with $\Pr[\mathsf{P}(g(X,R)) = b(X,R)] \geq \frac{1}{2} + \delta$, where hereafter R and X are iid uniformly distributed over $\{0,1\}^n$
- We prove $\exists\ (\frac{n^2}{\delta^2} \cdot s')$-size Inv with $\Pr[\mathsf{Inv}(f(X)){=}X] \in \Omega(\delta^3/n)$.

Focusing on a good set

Claim 10

There exists set $\mathcal{S} \subseteq \{0,1\}^n$ with

1. $\frac{|\mathcal{S}|}{2^n} \geq \frac{\delta}{2}$, and
2. $\Pr\left[\mathsf{P}(f(x), R) = b(x, R)\right]\right] \geq \frac{1}{2} + \frac{\delta}{2}, \qquad \forall x \in S.$

Proof: Let $\mathcal{S} := \{x \in \{0,1\}^n \colon \Pr\left[\mathsf{P}(f(x), R) = b(x, R)\right] \geq \frac{1}{2} + \frac{\delta}{2}\}$.

$$\begin{aligned}\Pr[\mathsf{P}(g(X,R)) = b(X,R)] &\leq \Pr[X \notin \mathcal{S}] \cdot \left(\frac{1}{2} + \frac{\delta}{2}\right) + \Pr[X \in \mathcal{S}] \\ &\leq \left(\frac{1}{2} + \frac{\delta}{2}\right) + \Pr[X \in \mathcal{S}]. \qquad \square\end{aligned}$$

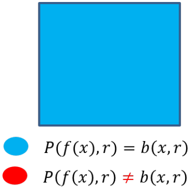
We conclude the theorem's proof showing that there exists a $\frac{n^2}{\delta^2}$-size Inv with

$$\Pr\left[\mathsf{Inv}(f(x)) = x\right] \in \Omega(\delta^2/n)$$

for every $x \in \mathcal{S}$. In the following we fix $x \in \mathcal{S}$.

The perfect case

$$\Pr\left[\mathrm{P}(f(x), R) = b(x, R)\right] = 1$$



In particular, $\mathrm{P}(f(x), e^i) = b(x, e^i)$ for every $i \in [n]$, for $e^i = (\underbrace{0, \ldots, 0}_{i-1}, 1, \underbrace{0, \ldots, 0}_{n-i})$.

Hence, $x_i = \langle x, e^i \rangle_2 = b(x, e^i) = \mathrm{P}(f(x), e^i)$

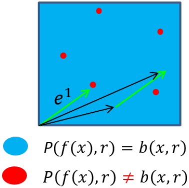
Algorithm 11 (Inverter Inv on input $y \in \mathrm{Im}(f)$)

Return $(\mathrm{P}(y, e^1), \ldots, \mathrm{P}(y, e^n))$.

$\mathrm{Inv}(f(x)) = x$.

Easy case

$\Pr\left[\mathrm{P}(f(x), R) = b(x, R)\right] \geq 1 - \frac{1}{4n}$



Fact 12

1. $b(x, w) \oplus b(x, y) = b(x, w \oplus y)$, *for every* $w, y \in \{0,1\}^n$.
2. $\forall r \in \{0,1\}^n$, *the rv* $(R \oplus r)$ *is uniformly distributed over* $\{0,1\}^n$.

Hence, $\forall i \in [n]$:

1. $x_i = b(x, e^i) = b(x, r) \oplus b(x, r \oplus e^i)$ for every $r \in \{0,1\}^n$
2. $\Pr\left[\mathrm{P}(f(x), R) = b(x, R) \wedge \mathrm{P}(f(x), R \oplus e^i) = b(x, R \oplus e^i)\right] \geq 1 - 2 \cdot \frac{1}{4n}$

Algorithm 13 (Inverter Inv on input y)

Return $(\mathrm{P}(y, R) \oplus \mathrm{P}(y, R \oplus e^1)), \ldots, \mathrm{P}(y, R) \oplus \mathrm{P}(y, R \oplus e^n))$.

$\Pr\left[\mathrm{Inv}(f(x)) = x\right] \geq 1 - 2n \cdot \frac{1}{4n} = \frac{1}{2}$

Proving Fact 12

1. For $w, y \in \{0,1\}^n$:

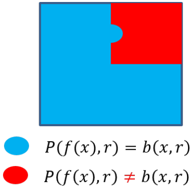
$$\begin{aligned} b(x,y) \oplus b(x,w) &= \left(\bigoplus_{i=1}^{n} x_i \cdot y_i\right) \oplus \left(\bigoplus_{i=1}^{n} x_i \cdot w_i\right) \\ &= \bigoplus_{i=1}^{n} x_i \cdot (y_i \oplus w_i) \\ &= b(x, y \oplus w) \end{aligned}$$

2. For $r, y \in \{0,1\}^n$:

$$\Pr\left[R \oplus r = y\right] = \Pr\left[R = y \oplus r\right] = 2^{-n}$$

Intermediate case

$\Pr\left[\mathsf{P}(f(x), R) = b(x, R)\right] \geq \frac{3}{4} + \frac{\delta}{2}$



For any $i \in [n]$

$$\begin{aligned}
&\Pr[\mathsf{P}(f(x), R) \oplus \mathsf{P}(f(x), R \oplus e^i) = x_i] \\
&\geq \Pr[\mathsf{P}(f(x), R) = b(x, R) \wedge \mathsf{P}(f(x), R \oplus e^i) = b(x, R \oplus e^i)] \\
&\geq 1 - \left(1 - (\frac{3}{4} + \frac{\delta}{2})\right) - \left(1 - (\frac{3}{4} + \frac{\delta}{2})\right) = \frac{1}{2} + \delta
\end{aligned}$$

Algorithm 14 ($\mathsf{Inv}(y)$)

For every $i \in [n]$:

1. Sample $r^1, \ldots, r^v \in \{0,1\}^n$ uniformly at random
2. Let $m_i = \mathsf{maj}_{j \in [v]}\{(\mathsf{P}(y, r^j) \oplus \mathsf{P}(y, r^j \oplus e^i))\}$

Output $(m_1, \ldots, m_n)$

Inv's success probability

The following claim holds for "large enough" v.

Claim 15

For every $i \in [n]$, it holds that $\Pr[m_i = x_i] \geq 1 - \frac{1}{2n}$.

Hence, $\Pr\left[\mathrm{Inv}(f(x)) = x\right] \geq \frac{1}{2}$. Proof: (of claim):

- For $j \in [v]$, let W^j be 1, iff $\mathsf{P}(f(x), r^j) \oplus \mathsf{P}(f(x), r^j \oplus e^i) = x_i$.
- We need to lowerbound $\Pr\left[\sum_{j=1}^{v} W^j > \frac{v}{2}\right]$.
- W^j are iids and $\mathrm{E}[W^j] \geq \frac{1}{2} + \delta$, for every $j \in [v]$

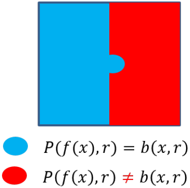
Lemma 16 (Hoeffding's inequality)

Let $X^1, \ldots, X^v$ be iids over $[0,1]$ with expectation μ. Then,
$\Pr\left[\left|\frac{\sum_{j=i}^{v} X^j}{v} - \mu\right| \geq \alpha\right] \leq 2 \cdot \exp(-2\alpha^2 v)$ *for every $\alpha > 0$.*

- Hence, the proof follows for $v = \left\lceil \log(n) \cdot \frac{1}{2\delta^2} \right\rceil + 1$.

The actual (hard) case

$\Pr[\mathsf{P}(f(x), R) = b(x, R)] \geq \frac{1}{2} + \frac{\delta}{2}$



- What goes wrong?
- $\Pr[\mathsf{P}(f(x), R) \oplus \mathsf{P}(f(x), R \oplus e^i) = x_i] \geq \delta$
- Hence, using a random guess does better than using P :-<
- Idea: guess the values of $\{b(x, r^1), \ldots, b(x, r^v)\}$ (instead of calling $\{\mathsf{P}(f(x), r^1), \ldots, \mathsf{P}(f(x), r^v)\}$)
- Problem: tiny success probability
- Solution: choose the samples in a correlated manner

Algorithm Inv

- For $\ell \in \mathbb{N}$ ($\approx \log \frac{n}{\delta}$, to be determined later), let $v = 2^\ell - 1$.
- In the following $\mathcal{L} \subseteq [\ell]$ stands for a non empty subset

Algorithm 17 (Inverter Inv on $y = f(x) \in \{0,1\}^n$)

1. Sample uniformly (and independently) $t^1, \ldots, t^\ell \in \{0,1\}^n$
2. Guess the value of $\{b(x, t^i)\}_{i \in [\ell]}$
3. For all $\mathcal{L} \subseteq [\ell]$: set $r^{\mathcal{L}} = \bigoplus_{i \in \mathcal{L}} t^i$ and compute $b(x, r^{\mathcal{L}}) = \bigoplus_{i \in \mathcal{L}} b(x, t^i)$.
4. For all $i \in [n]$, let $m_i = \text{maj}_{\mathcal{L} \subseteq [\ell]}\{\mathsf{P}(f(x), r^{\mathcal{L}} \oplus e^i) \oplus b(x, r^{\mathcal{L}})\}$
5. Output $(m_1, \ldots, m_n)$

- Fix $i \in [n]$, and let $W^{\mathcal{L}}$ be 1 iff $\mathsf{P}(f(x), r^{\mathcal{L}} \oplus e^i) \oplus b(x, r^{\mathcal{L}}) = x_i$.
- We need to lowerbound $\Pr\left[\sum_{\mathcal{L} \subseteq [\ell]} W^{\mathcal{L}} > \frac{v}{2}\right]$
- Problem: the $W^{\mathcal{L}}$'s are dependent!

Analyzing Inv's success probability

1. Let $T^1, \ldots, T^\ell$ be iid and uniform over $\{0,1\}^n$.
2. For $\mathcal{L} \subseteq [\ell]$, let $R^{\mathcal{L}} = \bigoplus_{i \in \mathcal{L}} T^i$.

Claim 18

1. $\forall \mathcal{L} \subseteq [\ell]$, $R^{\mathcal{L}}$ is uniformly distributed over $\{0,1\}^n$.
2. $\forall\, w, w' \in \{0,1\}^n$ and $\mathcal{L} \neq \mathcal{L}' \subseteq [\ell]$, it holds that $\Pr[R^{\mathcal{L}} = w \wedge R^{\mathcal{L}'} = w'] = \Pr[R^{\mathcal{L}} = w] \cdot \Pr[R^{\mathcal{L}'} = w'] = 2^{-2n}$.

Proof: (1) is clear. For (2), assume wlg. that $1 \in (\mathcal{L}' \setminus \mathcal{L})$.

$$\Pr[R^{\mathcal{L}} = w \wedge R^{\mathcal{L}'} = w']$$

$$= \sum_{(t^2,\ldots,t^\ell) \in \{0,1\}^{(\ell-1)n}} \Pr[(T^2,\ldots,T^\ell) = (t^2,\ldots,t^\ell)] \cdot \Pr[R^{\mathcal{L}} = w \wedge R^{\mathcal{L}'} = w' \mid (T^2,\ldots,T^\ell) = (t^2,\ldots,t^\ell)]$$

$$= \sum_{(t^2,\ldots,t^\ell):\ (\bigoplus_{i \in \mathcal{L}} t^i) = w} \Pr[(T^2,\ldots,T^\ell) = (t^2,\ldots,t^\ell)] \cdot \Pr[R^{\mathcal{L}'} = w' \mid (T^2,\ldots,T^\ell) = (t^2,\ldots,t^\ell)]$$

$$= \sum_{(t^2,\ldots,t^\ell):\ (\bigoplus_{i \in \mathcal{L}} t^i) = w} \Pr[(T^2,\ldots,T^\ell) = (t^2,\ldots,t^\ell)] \cdot 2^{-n}$$

$$= 2^{-n} \cdot 2^{-n} = \Pr[R^{\mathcal{L}} = w] \cdot \Pr[R^{\mathcal{L}'} = w'].\square$$

Pairwise independence variables

Definition 19 (pairwise independent random variables)

A sequence of rv's $X^1, \ldots, X^v$ is pairwise independent, if $\forall i \neq j \in [v]$ and $\forall a, b$, it holds that $\Pr[X^i = a \wedge X^j = b] = \Pr[X^i = a] \cdot \Pr[X^j = b]$.

- By Claim 18, $r^{\mathcal{L}}$ and $r^{\mathcal{L}'}$ (chosen by Inv) are pairwise independent for every $\mathcal{L} \neq \mathcal{L}' \subseteq [\ell]$.
- Hence, also $W^{\mathcal{L}}$ and $W^{\mathcal{L}'}$ are.
  (Recall, $W^{\mathcal{L}}$ is 1 iff $\mathsf{P}(f(x), r^{\mathcal{L}} \oplus e^i) \oplus b(x, r^{\mathcal{L}}) = x_i$)

Lemma 20 (Chebyshev's inequality)

Let $X^1, \ldots, X^v$ be pairwise-independent random variables with expectation μ and variance σ^2. Then, for every $\alpha > 0$: $\Pr\left[\left|\frac{\sum_{j=1}^{v} X^j}{v} - \mu\right| \geq \alpha\right] \leq \frac{\sigma^2}{\alpha^2 v}$.

Inv's success provability, cont.

- Assuming that Inv always guesses $\{b(x,t^i)\}$ correctly, then $\forall \mathcal{L} \subseteq [\ell]$:
  - $\mathrm{E}[W^{\mathcal{L}}] \geq \frac{1}{2} + \frac{\delta}{2}$
  - $\mathrm{V}(W^{\mathcal{L}}) := \mathrm{E}[(W^{\mathcal{L}})^2] - \mathrm{E}[W^{\mathcal{L}}]^2 \leq 1$
- Taking $v = 2n/\delta^2$ (hence $\ell = \lceil \log \frac{2n}{\delta^2} \rceil$), by Chebyshev's inequality for $i \in [n]$ it holds that

  $$\Pr[m_i = x_i] = \Pr\left[\frac{\sum_{\mathcal{L} \subseteq [\ell]} W^{\mathcal{L}}}{v} > \frac{1}{2}\right] \geq 1 - \frac{1}{2n}.$$
- By a union bound, Inv outputs x with probability $\frac{1}{2}$.
- Taking the guessing probability into account, yields that Inv outputs x with probability at least $2^{-\ell}/2 \in \Theta(\delta^2/n)$.
- Recalling that we guaranteed to work well on $\frac{\delta}{2}$ of the x's. We conclude that $\Pr[\mathrm{Inv}(f(x)) = x] \in \Theta(\delta^3/n)$.

Reflections

- Hardcore functions:

  Similar ideas allows to output $\log n$ “pseudorandom bits"

- Alternative proof for the leftover hash lemma:

  Let X be a rv with over $\{0,1\}^n$ with $H_\infty(X) \geq k$, and assume $SD((R, \langle R, X\rangle_2), (R, U)) > \alpha = 2^{-c \cdot k}$ for some universal $c > 0$.

$\implies$ $\exists$ (a possibly inefficient) D that distinguishes $(R, \langle R, X\rangle_2)$ from (R, U) with advantage α

$\implies$ $\exists$ P that predicts $\langle R, X\rangle_2$ given R with prob $\frac{1}{2} + \alpha$ (?)

$\implies$ (by GL) $\exists$ Inv that guesses X from nothing, with prob $\alpha^{O(1)} > 2^{-k}$

Reflections cont.

- List decoding:
  - Encoder $f\colon \{0,1\}^n \mapsto \{0,1\}^m$ and decoder g, such that for any $x \in \{0,1\}^n$ and c of hamming distance at most $(\frac{1}{2} - \delta)$ from $f(x)$: g examines $\mathrm{poly}(1/\delta)$ symbols of c and outputs a $\mathrm{poly}(1/\delta)$-size list that whp contains x
  - The code we used here is known as the Hadamard code
- LPN - learning parity with noise:

  Given polynomially many samples of the form $(R_i, \langle x, R_i \rangle_2 + \theta)$, for $R_i \leftarrow \{0,1\}^n$ and boolean $\theta_i \sim (\frac{1}{2} - \delta, \frac{1}{2} - \delta)$, find x.
- The difference comparing to Goldreich-Levin — no control over the R's.